

1/9

## HPV 52 L1 Nucleotide Sequence Alignment

52 L1 wt	( 1)	ATGTCCGTGTGGCGGCCTAGTGAGGCCACT <b>G</b> TGTACCTGCCTCCTGTACC
52 L1 R		.....C...A.A..ATCC..A..T....C...T....A..A...T..
52 L1 wt	( 51)	TGTCTCTAAGGTTGTAAGCACTGATGAGTA <b>T</b> GTGTCTCGACAAGCATCT
52 L1 R		A.....CTCT..C..C..A..C..C..CA.A..CTC.....
52 L1 wt	( 101)	ATTATTATGCAGGCAGTTCTCGATTACTAAC <b>C</b> AGTAGGACATCCCTATTT
52 L1 R		.C..C..C..T..TTCC...A....GT.G..T..C..T..C..A..C..C
52 L1 wt	( 151)	TCTATTAAAAACACCAGTAGTGGTAATGGT <b>A</b> AAAAGTTTAGTCCC
52 L1 R		....C..G.....TCCTCC.....C....G..G..C..G.....A..
52 L1 wt	( 201)	GGTGTCTGGCCTGCAATACAGGGTATTAG <b>A</b> ATTAAATTGCCGGACCC
52 L1 R		...C.....TT.....A..C..C.....C..G.....A.....A..
52 L1 wt	( 251)	ATAAATTGGTTTCCAGATA <del>C</del> ATCTTT <b>A</b> TAACCCAGAAACCAAAGG
52 L1 R		.C..G..C.....C.....C..TAG...C..C.....T.....A
52 L1 wt	( 301)	TTGGTGTGGCCTGTACAGGCTTGGAAATT <b>G</b> GTAGGGACAGCCTT
52 L1 R		....C.....T.....T..T.....C.....A..T..A..A..G..
52 L1 wt	( 351)	TGTGGTATTAGTGGGCATCCTTATTAAACAAGTTGATGATACTGAAA
52 L1 R		...C.....CTC...T..C..A..G..G.....C..C..C.....
52 L1 wt	( 401)	CCAGTAACAAATATGCTGGTAAACCTGGT <b>A</b> TAGATAATAGGAATGTTA
52 L1 R		..TC.....G..C.....G..A....C.....C..A.....G
52 L1 wt	( 451)	TCTATGGATTATAAGCAGACTCAGTTATGC <b>A</b> TTTAGGATGCAAACCTCC
52 L1 R		.....C..C.....A.....A..G..T..C..G..T..T..G..A..
52 L1 wt	( 501)	TATAGGTGAACATTGGGTAAGGGAACCC <b>T</b> TGTAATAATAATTCAAGGAA
52 L1 R		A..C.....C.....T..T.. <b>A</b> ....C..C..C..T..T..
52 L1 wt	( 551)	ATCCTGGGATTGTCCTCCCTACAGCTCA <b>T</b> TAACAGTGTAAACAGGAT
52 L1 R		.C..A..T..C.....A..AT.G..AT.G..C..TCC..C..C..A..C
52 L1 wt	( 601)	GGGGACATGGTAGATACAGGATTGGTGC <b>A</b> TGGATTAAACCTTGCA
52 L1 R		..T.....C..C..T..T..C.....T..C..C..C.....

FIG.1A

2/9

52 L1 wt	( 651)	AGCTAGTAAAAGT GATGTGCCATTGATATATGTAGCAGTGT ATGTAAGT
52 L1 R		....TC...GTCC..C..C..A..C..C..C...TC.TC... C.....
52 L1 wt	( 701)	ATCCAGATTATTGCAAATGGCTAGCGAGCCATATGGTGACA GTTTGTTC
52 L1 R		.C.....C..C.....TCT..A.....C.....TCC.....
52 L1 wt	( 751)	TTTTTCTTAGACGTGAGCAAATGTTGTTAGACACTTTT AATAGGGC
52 L1 R		..C..CT.G...A.A..A.....C..C.....C..C..C..A..
52 L1 wt	( 801)	CGGTACCTTAGGTGACCCTGTGCCAGGTGATTATATACAGGGTCTA
52 L1 R		T.....G.....A..T.....C..G..C..C....T..C.
52 L1 wt	( 851)	ACTCTGGCAATACTGCCACTGTACAAAGCAGTGCTTTTCTACTCCT
52 L1 R		.....T..C.....T.....C...TC.TC.....C..C..A.....A
52 L1 wt	( 901)	AGTGGTTCTATGGTAACCTCAGAACATTATTAAATAAA CCGTACTG
52 L1 R		TC.....C.....C.....G..C..C..G ..A.....
52 L1 wt	( 951)	GTTACAACGTGCGCAGGCCACAATAATGGCATATGTTGGGG CAATCAGT
52 L1 R		...G...A.A..T..A..T.....C..C..T..C.....T..C..A.
52 L1 wt	(1001)	TGTTGTACAGTTGGATACCACTCGTAGCACTAACATGA CTTTATGT
52 L1 R		....C.....C..C..C..T...A..ATCT.....C..G...
52 L1 wt	(1051)	GCTGAGGTTAAAAAGGAAAGCACATATAAAATGAAAATTAAAGGAATA
52 L1 R		.....A..C..G.....TC...C..C..G..C.....C..C .....
52 L1 wt	(1101)	CCTTCGTATGGCGAGGAATTGATTACAATTATTTCATTGTGCA
52 L1 R		.T.GA.A..C..T..A.....C..C..G.....C..C..C.....T.
52 L1 wt	(1151)	AAATTACATTAACAGCTGATGTTATGACATACATTATAAGA TGGATGCC
52 L1 R		.G..C..C..G..C.....C..C..T.....C..C.....C..T
52 L1 wt	(1201)	ACTATTAGAGGACTGGCAATTGGCCTTACCCACCACCG TCTGCATC
52 L1 R		....C..G..A.....C..TT.G..T.....A ..C..T..
52 L1 wt	(1251)	TTTGGAGGACACATACAGATTGTCACTTCTACTGCTATAAC TTGTCAA
52 L1 R		C.....A.....T.....C.....C.....C..C.....

FIG.1A

3/9

52 L1 wt	(1301)	AAAACACGCCACCTAAAGGAAAGGAAGATCCTTAAAGGACTATA TGTTT
52 L1 R		.G.....T.....A..G..T.....C..A..G.....C.....C
52 L1 wt	(1351)	TGGGAGGTGGATTAAAAGAAAAGTTTCTGCAGATTAGATCAG TTTCC
52 L1 R		.....A..C..C..G..G.....C.....T..C..G..C..A..C..
52 L1 wt	(1401)	TTTAGGTAGGAAGTTTGTTACAGGCAGGGCTACAGGCTAGGCC CAAAC
52 L1 R		A..G.....A.....C.....G..A..T..TT.G..A.....A..A..GT
52 L1 wt	(1451)	TAAAACGCCCTGCATCATCGGCCAACGTACCTCCACAAAGAAGA AAAAG
52 L1 R		.G..GA.A..A..TAGC..T..T...A.A..T.....C.....G...
52 L1 wt	(1501)	GTAAAAGGTAA (SEQ ID NO:3)
52 L1 R		..C..G..A...(SEQ ID NO:1)

FIG. 1C

4/9

## HPV 52 L1 R Nucleotide and Amino Acid Sequences

	M S V W R P S E A T V Y L P P V P
1	ATGTCCGTCT GGAGACCATC CGAACGCTACT GTCTACTTGC CACCAAGTTCC
	TACAGGCAGA CCTCTGGTAG GCTTCGATGA CAGATGAACG GTGGTCAAGG
	G S K V V S T D E Y V S R T S I Y
51	AGTCTCTAAG GTTGTCTCTA CCGACGAATA CGTCTCCAGA ACCTCCATCT
	TCAGAGATTG CAACAGAGAT GGCTGCTTAT GCAGAGGTCT TGGAGGTAGA
	Y Y A G S S R L L T V G H P Y F
101	ACTACTACGC TGGTTCCCTCT AGATTGTTGA CTGTCGGTCA CCCATACTTC
	TGATGATGCG ACCAAGGAGA TCTAACAACT GACAGCCAGT GGGTATGAAG
	S I K N T S S G N G K K V L V P K
151	TCTATCAAGA ACACCTCCTC CGGTAACGGT AAGAAGGTCT TGGTTCCAAA
	AGATAGTTCT TGTGGAGGAG GCCATTGCCA TTCTTCCAGA ACCAAGGTTT
	V S G L Q Y R V F R I K L P D P N
201	GGTCTCTGGT TTGCAATACA GAGTCTTCAG AATCAAGTTG CCAGACCCAA
	CCAGAGACCA AACGTTATGT CTCAGAAGTC TTAGTTCAAC GGTCTGGTT
	K F G F P D T S F Y N P E T Q R
251	ACAAGTTCGG TTTCCCAGAC ACTAGTTTCT ACAACCCAGA AACTCAAAGA
	TGTTCAAGCC AAAGGGTCTG TGATCAAAGA TGTTGGGTCT TTGAGTTCT
	L V W A C T G L E I G R G Q P L G
301	TTGGTCTGGG CTTGTACTGG TTTGGAAATC GGTAGAGGTC AACCATGGG
	AACCAGACCC GAACATGACC AACCTTTAG CCATCTCCAG TTGGTAACCC
	V G I S G H P L L N K F D D T E T
351	TGTCGGTATC TCTGGTCACC CATTGTTGAA CAAGTCGAC GACACTGAAA
	ACAGCCATAG AGACCAGTGG GTAACAACCT GTTCAAGCTG CTGTGACTTT
	S N K Y A G K P G I D N R E C L
401	CCTCTAACAA GTACGCTGGT AAGCCAGGTA TCGATAACAG AGAATGTTG
	GGAGATTGTT CATGCGACCA TTCGGTCCAT AGCTATTGTC TCTTACAAAC
	S M D Y K Q T Q L C I L G C K P P
451	TCTATGGACT ACAAGCAAAC TCAATTGTGT ATCTTGGGT GTAGGCCACC
	AGATACCTGA TGTCGTTTG AGTTAACACA TAGAACCCAA CATTGGTGG
	I G E H W G K G T P C N N N S G N
501	AATCGGTGAA CACTGGGGTA AGGGTACTCC ATGTAACAAC AACTCTGGTA
	TTAGCCACTT GTGACCCCCAT TCCCATGAGG TACATTGTTG TTGAGACCAT
	P G D C P P L Q L I N S V I Q D
551	ACCCAGGTGA CTGTCCACCA TTGCAATTGA TCAACTCCGT CATCCAAGAC
	TGGGTCCACT GACAGGTGGT AACGTTAACT AGTTGAGGCA GTAGGTTCTG
	G D M V D T G F G C M D F N T L Q
601	GGTGACATGG TCGACACTGG TTTCGGTTGT ATGGACTTCA ACACCTTGCA
	CCACTGTACC AGCTGTGACC AAAGCCAACA TACCTGAAGT TGTGGAACGT

FIG.2A

A S K S D V P I D I C S S V C K Y  
 651 AGCTTCTAAG TCCGACGTCC CAATCGACAT CTGTTCTCT GTCTGTAAGT  
       TCGAAGATTG AGGCTGCAGG GTTAGCTGTA GACAAGGAGA CAGACATTCA  
       P D Y L Q M A S E P Y G D S L F  
 701 ACCCAGACTA CTTGCAAATG GCTTCTGAAC CATA CGGTGA CTCC TTGTTG  
       TGGGTCTGAT GAACGTTTAC CGAAGACTTG GTATGCCACT GAGGAACAAG  
       F F L R R E Q M F V R H F F N R A  
 751 TTCTTCTTGA GAAGAGAACAAATGTTCTCGTC AGACACTTCT TCAACAGAGC  
       AAGAAGAACT CTTCTCTTGT TTACAAGCAG TCTGTGAAGA AGTTGTCTCG  
       G T L G D P V P G D L Y I Q G S N  
 801 TGGTACCTTG GGTGACCCAG TTCCAGGTGA CTTGTACATC CAAGGTTCCA  
       ACCATGGAAC CCACTGGGTC AAGGTCCACT GAACATGTAG GTTCCAAGGT  
       S G N T A T V Q S S A F F P T P  
 851 ACTCTGGTAA CACTGCTACT GTCCAATCCT CTGCTTTCTT CCCAACTCCA  
       TGAGACCATT GTGACGATGA CAGGTTAGGA GACGAAAGAA GGGTTGAGGT  
       S G S M V T S E S Q L F N K P Y W  
 901 TCTGGTTCCA TGGTCACCTC CGAATCCCAA TTGTTCAACA AGCCATACTG  
       AGACCAAGGT ACCAGTGGAG GCTTAGGGTT AACAAAGTTGT TCGGTATGAC  
       L Q R A Q G H N N G I C W G N Q L  
 951 GTTGCAAAGA GCTCAAGGTC ACAACAAACGG TATCTGTTGG GGTAACCAAT  
       CAACGTTCTC CGAGTTCCAG TGTTGTTGCC ATAGACAACC CCATTGGTTA  
       F V T V V D T T R S T N M T L C  
 1001 TGTTCGTCAC CGTCGTCGAC ACTACTAGAT CTACTAACAT GACCTTGTGT  
       ACAAGCAGTG GCAGCAGCTG TGATGATCTA GATGATTGTA CTGGAACACA  
       A E V K K E S T Y K N E N F K E Y  
 1051 GCTGAAGTCA AGAAGGAATC CACCTACAAG AACGAAAAC TCAAGGAATA  
       CGACTTCAGT TCTTCCTTAG GTGGATGTTTC TTGCTTTGA AGTTCCTTAT  
       L R H G E E F D L Q F I F Q L C K  
 1101 CTTGAGACAC GGTGAAGAAT TCGACTTGCA ATTCACTCTC CAATTGTGTA  
       GAACTCTGTG CCACCTCTTA AGCTGAACGT TAAGTAGAAG GTTAACACAT  
       I T L T A D V M T Y I H K M D A  
 1151 AGATCACCTT GACCGCTGAC GTCATGACTT ACATCCACAA GATGGACGCT  
       TCTAGTGGAA CTGGCGACTG CAGTACTGAA TGTAGGTGTT CTACCTGCGA  
       T I L E D W Q F G L T P P P S A S  
 1201 ACTATCTTGG AAGACTGGCA ATTGGTTTG ACTCCACAC CATCCGCTTC  
       TGATAGAAC TTCTGACCGT TAAGCCAAAC TGAGGTGGTG GTAGGGCGAAG  
       L E D T Y R F V T S T A I T C Q K  
 1251 CTTGGAAGAC ACTTACAGAT TCGTCACCTC CACTGCTATC ACCTGTCAA  
       GAACCTTCTG TGAATGTCTA AGCAGTGAAG GTGACGATAG TGGACAGTTT

6/9

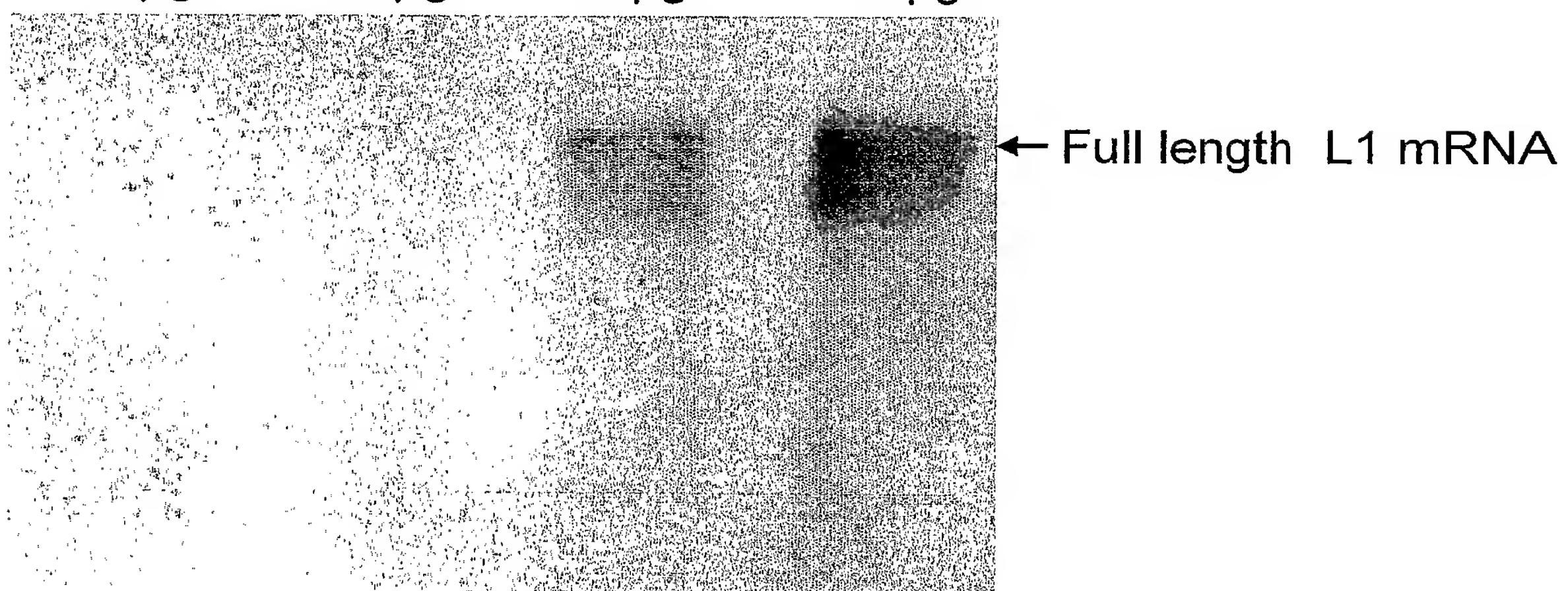
	N	T	P	P	K	G	K	E	D	P	L	K	D	Y	M	F	
1301	AGAACACTCC	ACCAAAGGGT	AAGGAAGACC	CATTGAAGGA	CTACATGTT	<b>C</b>	TCTTGTGAGG	TGGTTTCCA	TTCCTTCTGG	GTAACCTCCT	GATGTACAAG						
	W	E	V	D	L	K	E	K	F	S	A	D	L	D	Q	F	P
1351	TGGGAAGTCG	ACTTGAAGGA	AAAGTTCTCT	GCTGACTTGG	ACCAATTCCC		ACCCTTCAGC	TGAACCTCCT	TTTCAAGAGA	CGACTGAACC	TGGTTAAGGG						
	L	G	R	K	F	L	L	Q	A	G	L	Q	A	R	P	K	L
1401	ATTGGGTAGA	AAGTTCTTGT	TGCAAGCTGG	TTTGCAAGCT	AGACCAAAG	<b>T</b>	TAACCCATCT	TTCAAGAAC	ACGTTCGACC	AAACGTTCGA	TCTGGTTTCA						
	K	R	P	A	S	S	A	P	R	T	S	T	K	K	K	K	
1451	TGAAGAGACC	AGCTAGCTCT	GCTCCAAGAA	CTTCCACCAA	GAAGAAGAAG		ACTTCTCTGG	TCGATCGAGA	CGAGGTTCTT	GAAGGTGGTT	CTTCTTCTTC						
	V	K	R	*	(SEQ ID NO:2)												
1501	GTCAAGAGAT	AA	(SEQ ID NO:1)				CAGTTCTCTA	TT	(SEQ ID NO:7)								

FIG.2C

7/9

**Expression of HPV 52 L1 wt and 52 L1 R Transcripts**

52wt	52wt	52R	52R
5 µg	10 µg	5 µg	10 µg

**FIG.3**

8/9

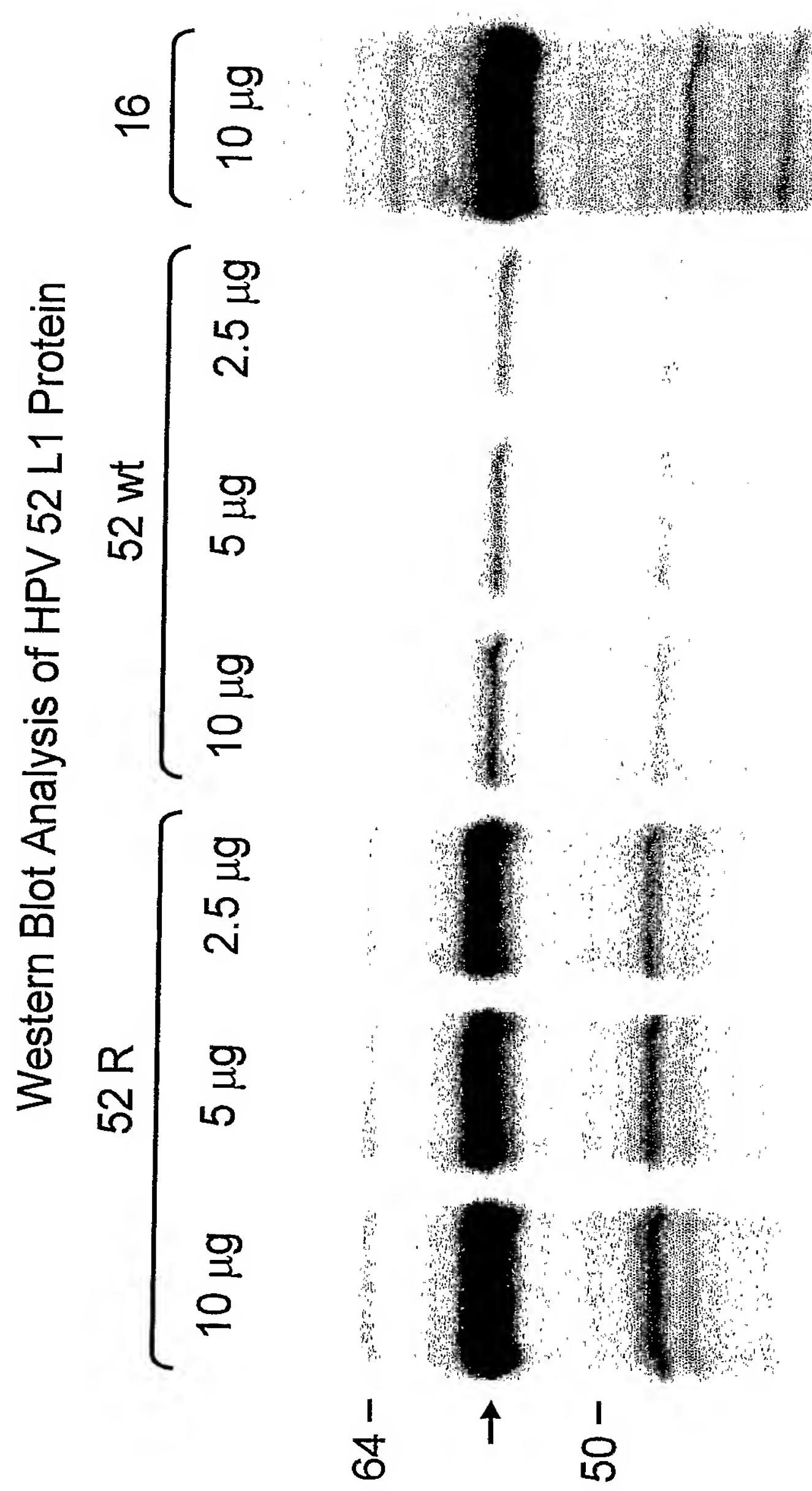


FIG. 4

9/9

Transmission EM of VLPs Composed of HPV 52 L1 R Protein Molecules

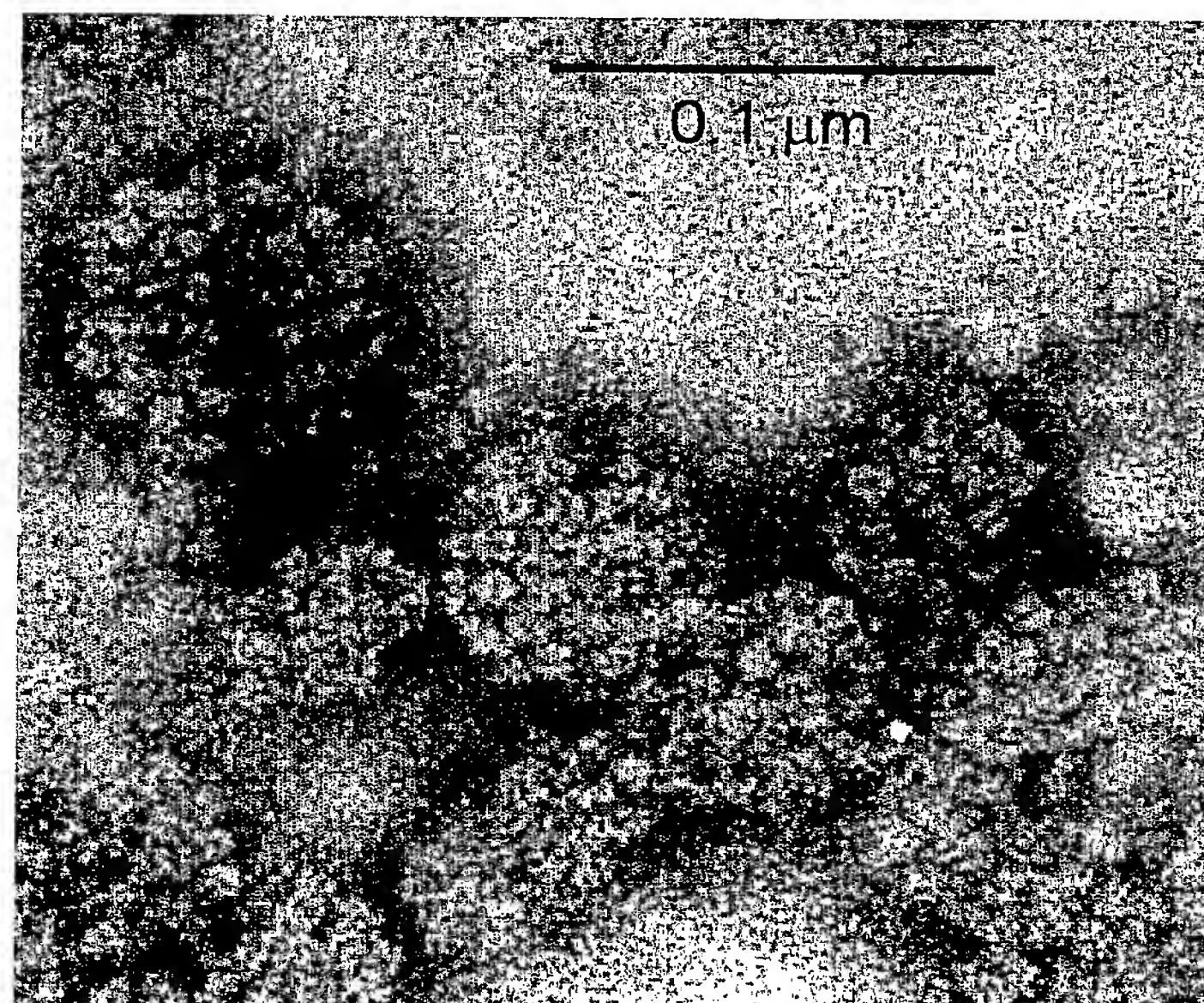


FIG.5